ISSN: 2182-2069 / E-ISSN: 2182-2077

Leveraging Machine Learning to Optimize Crop Yield Prediction: Insights from Sustainable Agriculture in Jordan

Muneer Nusir[1*], Mohammad Alshirah[2], Sahar Al Mashaqbeh[3], and Rayeh Alghsoon[4]

[1*]Prince Sattam Bin Abdulaziz University, Alkharj, Riyadh, Saudi Arabia.
moneer.techno@gmail.com, https://orcid.org/0000-0001-7938-741X

[2]Al al-Bayt University, Mafraq, Jordan. alshirah@aabu.edu.jo
https://orcid.org/0000-0001-6847-7938

[3]The Hashemite University, Zarqa, Jordan. saharm_al@hu.edu.jo
https://orcid.org/0000-0002-8939-2981

[4]Amman Al-Ahliyya University, Amman, Jordan. r.alghsoon@ammanu.edu.jo
https://orcid.org/0009-0006-5494-9061

Received: September 27, 2025; Revised: November 03, 2025; Accepted: December 26, 2025; Published: February 27, 2026

Abstract

Crop yield prediction is one of the crucial steps in food safety. Prediction with low error rates can support optimizing resources and facilitate informed decision making. Machine learning (ML) has been proven as useful tool for processing large agricultural datasets to forecast and predict the crop yield and other aspects of agriculture. ML models can facilitate informed decision-making and improve agricultural productivity. There have been no significant studies available on Jordan's data; this study utilizes the agricultural dataset from the Department of Statistics in Jordan and the climate dataset from the Climate Change Knowledge Portal to understand and build machine learning models that predict crop yield. This study aims to train and evaluate machine learning models, including regression, time series methods, Auto-Regressive Integrated Moving Average (ARIMA), and the Holt-Winters model for forecasting agricultural yields. Model's performance was evaluated with error metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), R-squared (R2), and Adjusted R-squared, etc. Results indicate Lasso regression and Random Forest are the most effective models with the lowest error rates and highest R-squared values. Lasso regression gets a R-squared value of 0.99 with an MAE of 0.104, while Random Forest achieves an R-squared value of 0.99 with an MAE of 0.108 across all evaluation metrics.

Keywords: Lasso Regression, Random Forest Regression, ARIMA, Holt-Winters, Crop Yield Prediction, Smart Agriculture, Machine Learning Models.

1 Introduction

Developing countries are strengthening economically when compared to advanced economies. This economic growth and prosperity increased the demand for food supply. The demand for cereals, animals, dairy, fruits, and vegetables is on the rise. Intrinsic demand. To meet the increased demand, the agricultural sector must enhance and adopt modern technologies while also overcoming the challenges

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 1 (February-2026), pp. 398-417.
DOI: *10.58346/JISIS.2026.I1.023*

*Corresponding author: Prince Sattam Bin Abdulaziz University, Alkharj, Riyadh, Saudi Arabia.

of climate change, depleting resources, and changing dietary patterns. Adopting machine learning is inevitable in forecasting the crop yield and weather patterns.

To incorporate sustainable practices and optimal utilization of resources and to mitigate the factors affecting the environment is a priority in agricultural practices. Achieving food security with smart agricultural practices and prediction using machine learning is highly necessary (Paudel et al., 2021; Hukare & Kumbhar, 2025). In smart agriculture, IoT is one of the main technologies, which enhances all phases of agriculture, from soil preparation and pest control to logistical support. IoT is data-centric, collecting real-time data for analysis using machine learning models for informed decision-making. Implementing IoT comes with various challenges; overcoming these challenges associated with IoT is essential for leveraging smart agriculture technologies. Thus, it is crucial to understand the future trends of IoT in agriculture and the research gaps in successfully implementing these state-of-the-art technologies (Ayaz et al., 2019).

This study introduces a novel machine learning approach tailored to enhance crop yield forecasting to support the sustainable growth of wheat crop yields in Jordan. Jordan is still in the early stages of climate-smart agriculture production (Elavarasan & Vincent, 2020; Singh et al., 2021). Hence, this study has important ramifications and implications for smart agriculture and farming practices in the region. This study's use of ML algorithms is considered a particularly intriguing mechanism. In comparison to traditional approaches, this study offers a clearer and unique solution in forecasting crop yields (Albishi et al., 2017; Samal et al., 2022).

Data collection is one of the early significant stages of this study; data gathered here is from the crop production division of the Department of Statistics, Jordan, which in turn collected it through surveys with farmers. Data collection using IoT and smart agricultural technologies is more convenient and challenging. Some of the challenges are data security, privacy, integrity, and maintaining secure communication channels (Samal et al., 2022). To overcome these challenges and establish secure communication that maintains privacy and security, it is mandatory to implement established privacy and security protocols (Albishi et al., 2017). Privacy-enhancing technologies are categorized into seven groups: personal data protection, enforcement mechanisms, data control, anonymization, selective data disclosure, authorization, and comprehensive privacy-preserving strategies (Cha et al., 2018). Requirements of secure IoT systems are data integrity, privacy, authentication mechanism, access control, and non-repudiation (Ferrag et al., 2017).

IoT, along with machine learning models, collects and analyzes large amounts of data for informed decision-making during plantation, optimization of water usage, harvesting for better crop yield, and achieving minimal wastage of resources and labor. Thus, smart agriculture technologies became a cornerstone in modern agricultural practices, allowing different agricultural systems to grow and meet the demands of a growing population (Ferrag et al., 2017; Elbasi et al., 2023). IoT comes with many advantages and challenges. IoT devices produce poisonous waste and have an impact on the environment and human well-being. To solve these problems, these systems need to be used in a way that is both effective and environmentally friendly. The solution is to integrate green IoT for sustainable agricultural practices. Machine learning (ML) makes no assumptions about the functional form, probability distribution, or smoothness of the data model, unlike classic statistical approaches (Samal et al., 2022). According to Elavarasan & Vincent (Elavarasan & Vincent, 2020), ML approaches can ascertain the correlation between independent and dependent variables through data analysis. ML approaches rely on structures that are nonparametric and semiparametric, and their validity is determined by precision in predictions. If the class and other properties fulfill certain probability distributions, a nonparametric technique does not need any previous presumptions on those distributions' shape (Albishi et al., 2017).

A proposed model is developed to efficiently predict the harvest production by conserving the valid data distribution with a precision of 93.7% using a deep learning algorithm.

Adaptive Forecasting Models: Techniques like ARIMA and Holt-Winters models have been used for time-series yield prediction, but their effectiveness is limited by model assumptions and data variability. However, despite these improvements in applications, the available models struggle with generalizability across different regions, fail to integrate multiple environmental and soil-related factors, and often lack transparency in their decision-making processes. Accordingly, our research presents a novel machine-learning approach for crop yield forecasting in Jordan, where climate-smart agricultural practices are still in early development. In this research, several regression models for predicting the yield of crops like wheat, cotton, and lentils are applied depending on soil, weather, and crop parameters. ML methods are used to train the models. To predict crop yield, linear regression models, lasso regression, random forest regression, XGBoost, SVM regression, decision tree analysis, ridge regression, Elastic Net regression, and polynomial regression are utilized in this research study. In this study ARIMA and Holt-Winter methods are used for adaptive forecasting. The aims of this research are as follows:

- Improved Prediction R-squared value by integrating various machine learning methods such as regressions, XGBoost.
- To build a robust predictive model by incorporating crop parameters like planted area, harvested area and produced quantity.
- Measuring effectiveness of proposed approach with comparative analysis of multiple forecasting methods.

This study achieved a significant and clear machine learning-based solution for the prediction of crop yield by identifying the gaps in existing prediction models. Further literature review explains the research gaps in these models. This study adopted multiple machine learning algorithms and measured the R-squared value and error metrics to identify the best-performing models in crop yield prediction. Further crop yield forecasting is done for the next 10 years, from 2023 to 2032, by using adaptive forecasting methods: ARIMA and the Holt-Winters model. Findings from this study are discussed in detail in the results and discussion sections.

Literature Review

Effectiveness of Factors Influencing Crop Yield

The main factors that influence the crop yield fall under three categories: technological, biological, and environmental. To overcome the challenges posed by these factors, adopting smart agriculture to achieve sustainable goals of agriculture is inevitable (Liliane & Charles, 2020). Technical factors that impact the crop yield are agricultural practices, management, and decision-making. Data collected from sensors, satellites, and other sources can help farmers make informed decisions. Integrating cutting-edge technologies like IoT, AI, and Machine Learning can give farmers an advantage in soil management, water management, pest control, and better forecasting. (Xu et al., 2022) Reviewed the agricultural IoT technology to summarize the technologies of agricultural IoT and further identify the problems in agricultural IoT.

Droughts are accelerated due to climate change. This environmental factor has significantly impacted the crop yield. The study by Mukherjee et al. (Mukherjee et al., 2018) investigated how variation in land use practices contributed to the escalation of droughts. Rainfall patterns affect the crop yield significantly due to changes in the soil properties. Lad et al., 2022 studied the crop yield in India between 2000 and 2015 and analyzed the data by integrating supervised machine learning algorithms to recommend different crops to plant based on the levels of nitrogen, phosphorus, potassium, humidity, pH levels, and temperature.

Machine Learning Techniques for Predicting Crop Yield

Forecasting crop yield is significant for farmers to make decisions on resource allocation, crop production management, and logistical strategies. Several regression models have proven to perform better in predicting the crop yield based on the dataset attributes, interpretability, and computational efficiency. The Random Forest regression model has shown a significant R-squared value and a low error score of R2. Further combination of regression models with other machine learning techniques results in low R2 and better crop yield prediction models in agriculture (Jorvekar et al., 2024). Datasets with time series combined with other attributes like soil properties, climate data, and satellite images give a more accurate picture of the agricultural yields. Satellite image data of Sanliurfa, Turkey, for a wheat yield prediction model with four spectral indices: Normalized Difference Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Green Normalized Difference Vegetation Index (GNDVI), and Modified Soil-adjusted Vegetation Index (MSAVI), using a linear regression model, shows the highest correlation between the flowering stage and actual yield values. A limitation of this type of study is the resolution of the satellite images; with better resolution of images, a better prediction model can be achieved (Kaya & Polat, 2023).

Incorporating machine learning technologies to accurately predict crop yield can help farmers in determining possibilities and threats. Supervised learning regression models are trained with the dataset obtained from the open data portal of Telangana, India. Results show that the Regression Model outperformed other models in predicting the crop yield (Panigrahi et al., 2023). Along with historical data, other attributes, such as weather, soil properties, and fertilizer usage, enable the development of robust predictive models that assist farmers and policymakers in making better decisions on resource allocation and crop management. Machine learning models trained with the Random Forest algorithm performed better in comparison with other models (Gladence et al., 2021). In one more study, when it comes to accurate large-scale estimates of actual yields of wheat in Australia, support vector regression with radial basis function came out as the best learner, and this study shows that machine-learning regression methods applied with time series, climate, and satellite image data can achieve better crop yield prediction across years (Kamir et al., 2020). Forecasting models with advanced machine learning models other than least squares regression models are studied using the boosted regression technique (BRT). Climate and crop yield data result in BRT's capability of using climate and crop yield data to better predict and also identify the interactions between different variables that affect crop yields (Sidhu et al., 2023).

In one study, several regression-based learner models and two ensemble models were tested for crop yield prediction, but Support Vector Regression achieved the highest learning efficiency. Ensemble models are high in computing cost, yet they have not significantly enhanced the R-squared value. Additionally, increasing the amount of training data improved performance across all models. In another study, the effectiveness of machine learning techniques for predicting corn yield in Iowa State was examined using remote sensing data. The results indicated that machine learning methods are effective for yield estimation, with Deep Learning (DL) delivering particularly stable and favorable outcomes (Meshram et al., 2021). Further, several machine learning models were assessed and analyzed based on their performance using the R-squared error metric. "Based on their performance, we utilize the R-squared error metric. Among the evaluated models, Random Forest combined with the Forward Feature Selection algorithm revealed the best performance (Nusir, 2024).

In this paper, several regression models are performed and trained. The proposed research methodology to develop the model is explained in the following section.

2 Materials and Methods

Due to the complexity of machine learning models for agricultural-based studies, creating the machine learning models is quite a tedious and complex task. The modeling problem can be reflected as either a classification or a regression-based problem (Nusir, 2024). This study uses the agriculture dataset from the Department of Statistics in Jordan (Pargent et al., 2022) and weather data (Páez & Boisjoly, 2023) to train the machine learning models, which can predict the yield of crops. Further details of data collection are given in section 3.1. Agricultural dataset usage and ethical issues surrounding it are addressed by researchers. Machine learning models with small datasets cannot give the same R-squared value as big data; inaccurate data and recommendations may mislead the farmers, resulting in loss of yield and business and even impact the environment (Zhang et al., 2014). Data ownership, where farmers are concerned about sharing the data because of many speculations about regulatory enforcement and fee impositions, and it could be used by traders to lower the prices of yield (Mark, 2019). This study compares the traditional and advanced machine learning models to predict the crop yield. Machine learning models are trained and tested using the dataset specific to Jordan. The results of the performance of these models are discussed to understand the suitable models and future development.

This experiment is performed on a Windows platform running on a Lenovo desktop with an i7 processor and 16 GB of RAM. The Python programming language on Jupyter IDE is used to train and build the models. The Python version used in this study is 3.11.5. A total of ten models were built to predict the crop yield. The algorithms selected here are regression algorithms, from a simple multi-linear algorithm to an ensemble XGBoost algorithm and ARIMA time series forecasting, which are discussed in the next section.

These datasets from the websites of the Department of Statistics, Jordan, and the Climate Change Knowledge Portal include 5 crop details from 1999 to 2022. There is a total of six columns: year, average yield, harvested area, planted area, crop type, and production. Production is the target variable; the prediction of crop production or yield depends on the remaining 5 attributes in the dataset. Data from the second source contains the temperature mean (Tmean, ◦C), maximum temperature Tmax in ◦C), minimum temperature Tmin in ◦C), and precipitation (Prcp, mm) of Jordan from 1999 to 2022. These two datasets are merged into a single dataframe using the panda's library in Python. The description of the dataset is presented in Table 1. The dataset is further preprocessed and cleaned to check for any missing values, zero variance, and outliers. Section 5 covers the preprocessing details.

The cleaned dataset is used for exploratory data analysis (EDA), and Section 6 covers the EDA results. This dataset consists of data from 1999 to 2022; the crop type column is a categorical data column, which is converted into an indicator variable with the get_dummies function. Planted area and harvested area are the number of acres each year from 1999 to 2022, average yield represents the yield or production from harvested, and production is presented in metric tons. Temperature is in centigrade, and precipitation is in millimeters.

Machine learning models can be leveraged in agriculture for early detection of crop disease identification, crop yield prediction, weather forecasting, crop price prediction, and species identification (Zhang et al., 2014). Results from machine learning models are vital information for farmers in informed decision-making at each step in agriculture. The general steps followed by farmers are as follows: The first step is the selection of crops. The second step is regarding the preparation of the land, the third step is seed sowing, and the fourth step is about irrigation & fertilizing. After that, the crop maintenance step is started, then the harvesting step begins, and finally, the post-harvesting activities start (Mark, 2019). The executed methodology for this research is described in Figure 1. Step 1 is data collection from multiple sources, step 2 is data preprocessing where data is cleaned by removing null values and outliers, step 3 is exploratory data analysis, in step 4 the data is split into test data and training data, step 5 is training the machine learning model of the training dataset, and step 6 is result analysis after testing the model in the test dataset.

The following steps explain the process of collecting data, processing, exploring, splitting data, and training data.

Data Collection

First step: data is collected from the websites of the Department of Statistics, Jordan, and the climate change knowledge portal. Data is downloaded in comma-separated value (CSV) format. The Department of Statistics provides data and reports on the socioeconomic aspects of Jordan, including the environment, agriculture, and much more. Datasets encompass the following features: Historical crop yield data from 1999 to 2022 from Jordan for the crop's wheat, barley, chickpea, lentil, and vetch. Climate variables: minimum temperature, maximum temperature, mean temperature, and precipitation. Datasets from two sources are merged into a single dataframe to train and test the models. After merging, there are a total of 10 attributes, of which production is the target variable, and the remaining 9 variables are independent variables used to predict production. Correlation of the attributes is presented in the data exploratory data analysis phase of the study.

Data Processing

Preprocessing of data after collection. Python programming is used in this study, and Jupyter Notebook is the development environment. The files, including EDA and machine learning models, are uploaded to GitHub for public access (Hastie, 2009). During data preprocessing, the dataset is loaded as a panda Data Frame. The first step is to check for any null values and zero-variance columns. This dataset has no null values and zero variance. The crop column is a string type with the name of the crop, which is encoded from string to binary for machine learning. The one-hot encoding technique is applied to convert this column into binary code for each crop, which makes it efficient for machine learning models, which results in the addition of new columns separate for each crop: crop_Wheat, crop_Barley, crop_Chickpea, crop_Lentil, and crop_Vetch. With the addition of these columns, the total number of columns increased from 10 to 13. Now the dataset has 12 independent variables and one dependent variable, that is, production. Encoding allows the machine learning algorithm to perform better (Basso & Liu, 2019). The planted area and harvested area columns represent acres of land where a crop is planted and harvested, and the production column presents metric tons of crop yield. These values are large floating-point numbers. Large numbers affect the machine learning model's performance; thus, these columns are transformed with the smart transform and log transform. Smart transform is applied on independent variables, and log transformation on the dependent variable 'Production.' After preprocessing the data, the next step is to

conduct exploratory data analysis (EDA).

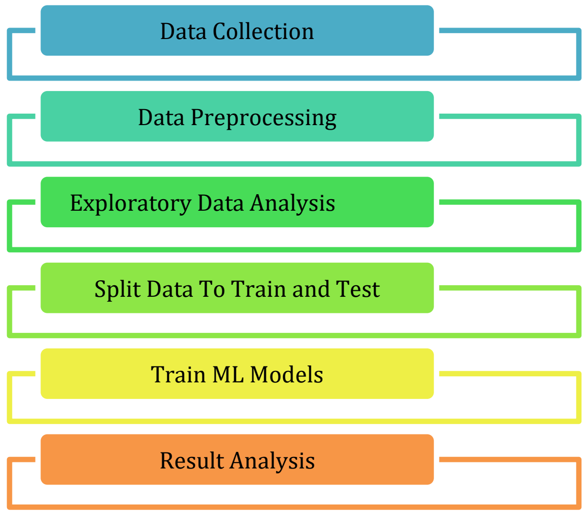

Figure 1: Proposed research methodology for crop yield prediction

Table 1: Dataset description

	Column name	**Datatype**	**Values**
1	Year	Date	Year
2	Planted area	Floating point	Dunum or Acre
3	Harvested area	Floating point	Dunum or Acre
4	Average yield	Floating point	Metric Tons per Dunum
5	Production	Floating point	Metric Tons
6	Crop	String	Wheat, barley, lentil, chickpea, vetch
7	Temperature Minimum	Floating point	Minimum temperature of the year
8	Temperature Maximum	Floating point	Maximum temperature of the year
9	Temperature Mean	Floating point	The mean temperature of the year
10	Precipitation	Floating point	Millimeter (mm)

Exploratory Data Analysis (EDA)

EDA presents the basic information, data interpretability, and visualization of the data to make it clear and simple to understand (Hastie, 2009). Production and average yield of each crop by year (1999-2022) can be visualized with the data of each crop in Figure 2.

The dataset here is of 13 columns, among which the 'Production' column is the target column, which is being predicted or measured using the machine learning models, and the remaining 12 columns are independent columns that are used in predicting and measuring the target column. All of the columns contribute to predicting the production by calculating the correlation between the independent and target variables. To find the correlation, Pearson's Correlation Coefficient method is applied, and the results are shown in Table 2. Pearson's correlation coefficient measures the strength of the linear relationship between 2 quantitative variables. If the value of this coefficient is close to 1 1, this indicates a strong linear relation between these two variables. If the value is close to 0, it indicates poor correlation. Eq. (1) represents Pearson's correlation coefficient. Harvested area and planted area columns contribute the most to predicting production.

$$r = (n * \Sigma xy \ \Sigma x * \Sigma y) / sqrt((n * \Sigma x^2 \ (\Sigma x)^2) * (n * \Sigma y^2 \ (\Sigma y)^2 \qquad (1)$$

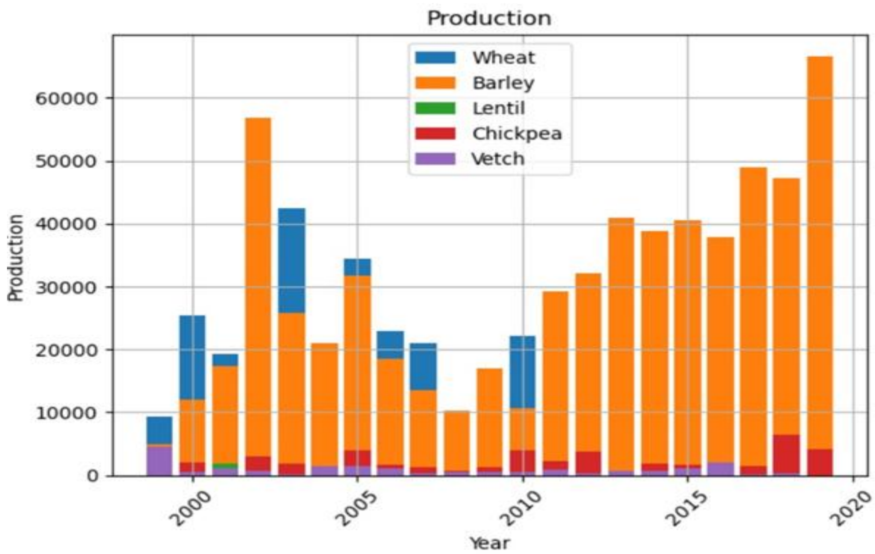

Figure 2: Yearly production of crops

Table 2: Correlation coefficients

Year	**0.162050**
Planted Area	0.732012
Harvested Area	0.935576
Temp_mean	0.050504
Temp_min	0.093477
Temp_max	0.007114
Precipitation	0.169525
Average Yield	0.117689
crop_Barley	0.634870
crop_Chickpea	0.284775
crop_Lentil	0.338342
crop_Vetch	0.325990
crop_Wheat	0.314238
Production	1.000000

Split Data to Train and Test

After preprocessing the data and before building the ML model, the data is split into data for training and testing the model, and data to test the model to measure the model's performance (Chen & Guestrin, 2016). Machine learning algorithms are used to train the models. The split dataset is categorized into four groups: Xtest, Xtrain, Ytest, and Ytrain. X_train and y_train are used to train the model. To test the model X_test given as input to the model it gives y_pred as output which is compared to Y_test to compare the R-squared value ad error rates (Behera et al., 2021). Model R-squared value is measured with R-squared value rate, error rate of each model is a measure to prediction error of models. All the models in this study are measured with same criteria to compare and evaluate the performance and error rate.

Training Machine Learning Models

Machine learning is highly effective in processing complex high-dimensional data. Agricultural datasets with nonlinear relationships among variables can be modeled with machine learning algorithms, which are effective in using agricultural, environmental, and climatic variables with nonlinear relationships. ML algorithms such as decision trees, random forests, and XGBoost outperform the normal statistical models. Random forest is a robust algorithm to handle high-dimensional data to predict chickpea yields, and Gaussian models predict wheat and pearl millet prediction with high R-squared value (Patil et al., 2025). In a study, deep learning models perform with an R-squared value in predicting crop yield using remote sensing imagery of drought maps and spatial data. This novel approach of an ensemble of deep learning algorithms, such as long short-term memory, convolutional neural networks, deep neural networks, and recurrent neural networks, and further genetic algorithms, is used to enhance the mapping of groundwater potential zones (Rahman et al., 2023). The support vector machine algorithm is used to prescribe the best crop, and deep learning models, long short-term memory, and recurrent neural networks are used to predict the crop yield. This model's performance is compared with the existing model (Agarwal & Tarar, 2021). Feature selection before building a machine learning model enhances the performance of the model. In the agricultural framework, feature selection is done with the relief algorithm and feature extraction with the linear discriminant analysis algorithm. Classification algorithms particle swarm optimization-support vector machine (PSO-SVM), K-nearest neighbor, and random forest show better R-squared value, sensitivity, and specificity (Gupta et al., 2022). This approach of integrating multiple machine learning algorithms along with feature selection and feature extraction provides better crop yield prediction, which enables informed decisions, supports sustainable agricultural practices, and promotes food security.

Multiple Linear Regression

Multiple linear regression establishes the linear relationship between multiple independent variables with dependent variable. It is a flexible and powerful model to quantify how several factors contribute to an outcome; errors in this model are normally distributed with constant variance. It is widely used in forecasting crop yields. MLR models the relationship between multiple independent variables and the dependent variable. Eq. (2) represents multiple linear regression, where $\hat{y}$ (yhat) represents the predicted value of the dependent variable based on independent variables, β_0 (betanaught) is the y-intercept, which represents the predicted value of dependent variable (y) when all the independent variables are zero, β_1 (betaone) to β_n (betan) are the regression coefficients for each independent variable (x_1 to x_n) and ε (epsilon) represents the error term.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (2)$$

Random Forest

Random Forest is an ensemble learning technique used for regression tasks, which functions by generating multiple decision trees during training and producing the average forecast of the individual trees. This method enhances the forecast R-squared value and mitigates overfitting. Each individual tree is created by random selection of instances from the training data. Variability in data selection overcomes the high correlation and builds the robust model. Predictions are derived by aggregating individual predictions and considering the voting majority among the individual trees. Random forests take advantage by combining multiple weak learners (individual decision trees) to achieve a robust and accurate prediction. Random forest is shown in equation 3; y_cap is the prediction, and N represents the number of trees. Ti(x) is the prediction of the ith tree. Greater tree density in a forest leads to more accurate predictions (Zou & Hastie, 2005; Patil et al., 2025). Random Forests is a nonparametric advanced classification and regression tree (CART) analysis method.

$$\hat{y} = \frac{1}{N} \sum_{i}^{n} T_i(x) \qquad (3)$$

Lasso Regression

Lasso regression integrates regularization and feature selection. It adds penalty term to linear regression and shrinks some coefficients to zero resulting in effectively removing features with less importance to deal with overfitting of model. Lasso regression handles high-dimensional data effectively. L1 regularization of Lasso regression add the penalty proportional to the coefficients value. The Lasso regression formula is represented in equation 4 (Chen & Guestrin, 2016)

$$minimize: 1/2 \ast \Sigma\ (y\_i\ \ \beta\_0\ \ \Sigma\beta\_j\ \ast\ Xin)\ \hat{}2\ +\ \lambda\ \ast\ \Sigma|\beta\_j| \quad (4)$$

Polynomial Regression

Polynomial regression is used to model nonlinear variables with curvilinear relationship using polynomial equation. Relationship between dependent and multiple independent variables is explained with curvilinear relation, it fits the data by predicting the best curve, it used when data points are not captured with linear regression. Model increases or decreases the degree till it finds the best possible model without over-fitting or under-fitting. Equation 5 shows the polynomial regression where xn is the degree of polynomial.

$$y\ =\ b0\ +\ b1 \ast x\ +\ b2 \ast x\hat{}2\ +\ b3 \ast x\hat{}3 + \ldots + bn \ast x\hat{}n\ +\ e \quad (5)$$

Ridge Regression

Ridge regression deals with the collinearity among independent variables. The formula for the ridge regression is equation 6 shows the λ (lambda) variable resolves the multicollinearity by adding penalty to shrink the coefficients. In Ridge regression L2 regularization adds penalty proportional to the square of the value of coefficients to shrink the coefficients nearly to zero.

$$\beta\ =\ (X\hat{}T\ X\ +\ \lambda I)\hat{}1\ X\hat{}T\ y \qquad (6)$$

Elastic Net Regression

Elastic Net regression is used to handle multicollinearity it uses the penalties of Ridge regression and Lasso regression to overcome the overfitting and increase the model R-squared value . The objective is to overcome the limits of each unique method while capitalizing on its advantages. As shown in equation 7, λ1 and λ2 are the regularization parameters to shrink the coefficients and α to balance the L1 and L2 penalties.

$$minimize: 1/2\ \ast\ \Sigma(y\_i\ \ \beta\_0\ \ \Sigma\beta\_j\ \ast\ x\_ij)\hat{}2 + \lambda 1 \ast \Sigma|\beta\_j| +\ \lambda 2/2\ \ast\ \Sigma\beta\_j\hat{}2 \quad (7)$$

Support Vector Machines (SVM)

Support Vector Machines (SVMs) are typically utilized for classification; however, in this case, they are being applied to regression problems. The aim is to identify a hyperplane that optimally accommodates the dataset while accounting for a specific margin of error. The initial component in the objective function shown in Eq. (8) represents the complexity of the model, whereas the next component quantifies the degree of error. The regularization parameter C determines the balance between accurately fitting the data and the model complexity. The parameter epsilon specifies the epsilon-tube within which the residuals in epsilon SVR are not subject to a penalty. In this study, default hyperparameters were applied to the SVM regressor; the regularization parameter C is 1.0, and the default kernel is Radial Basis Function (RBF).

$$minimize: 1/2\ ||w||\hat{}2\ +\ C\ \ast\ \Sigma(max(0, |y\_i\ \ w\hat{}T\ x\_i\ \ b|\ \ epsilon) \quad (8)$$

Decision Tree Regression

Decision tree regression is a nonparametric technique that builds a model like a tree to represent decisions and their potential outcomes, which includes chance events, costs of resources, and utility (Rifna et al., 2024). The algorithm initiates by identifying the optimal feature to divide the data. The data is separated into subsets according to the split. This procedure is iteratively performed for each subset, generating new nodes and branches until the end condition is satisfied. Although decision tree regression does not have a specific formula, its algorithmic methodology makes it a robust and easily understandable solution for a wide range of regression situations. In this study, the decision tree regressor uses default hyperparameters; the criterion is mse, the split method is 'best,' and maxdepth is none, which indicates the tree can grow indefinitely. CART includes multiple decision trees; it combines the predictions from different decision trees in the forest, as shown in Eq. 9 (Meshram et al., 2021).

$$Minimize: (1/2N)\ \Sigma\ (yi\ \ \hat{y}i)\ \hat{}2\ +\ \lambda\ \Sigma|\beta j| \qquad (9)$$

XGBoost

XGBoost, eXtreme Gradient Boosting algorithm is used for classification and regression. It is an ensemble learning technique. It applies gradient boosting to enhance the prediction in previous trees in the ensemble of decision trees in sequential order. This algorithm corrects the errors and enhance the prediction in the previous trees, by combining weak learners into a strong ensemble. Eq. (9) shows the concept of XGBoost. ŷ(x) is the predicted output for the data point (x), f_0(x) is an initial prediction, and Σ f_t(x) is the sum of predictions for all the decision trees in the ensemble. f_t(x_t) is the prediction output from the decision tree, and T is the total number of trees in the XGBoost model (Patil et al., 2025). It combines L1 and L2 regularization to reduce the overfitting and the correction is done by the gradient of the loss function to minimize the mean squared error between actual values and predicted values. Gradient boosting optimization corrects prior errors and build a strong and accurate prediction model.

$$\hat{y}(x)\ =\ \Sigma\ f\_t(x)\ =\ f\_0(x)\ +\ \Sigma\ f\_t(x\_t)\ where\ t\ =\ 1\ to\ T \qquad (10)$$

3 Adaptive Forecasting Methods

ARIMA

Autoregressive Integrated Moving Averages (ARIMA) is the time-series forecasting method, it combines autoregression modeling, differencing and moving averages to predict the future values. This model uses the historical data to identify the trend using autoregression. Trends are removed with differencing component I (Integrated) to make the data-points stationary, moving averages make the adjustments by dealing with past prediction errors. ARIMA(p,d,q) model components are p,d,q. Autoregression is 'p' which uses historical data, Integrated 'd' removes the trend to make time-series data stationary, 'q' is the moving averages.

Holt-Winter

The Holt-Winter model is another time-series forecasting method, which deals with trend and seasonality in the data. This model is suitable for agricultural datasets, as it shows a trend and seasonality in production and temperature. The Holt-Winter model has three components: level is the baseline value of time-series data, trend presents the upward or downward trajectory of the data points, and seasonality shows the repeated patterns in the data. Holt-Winter is applied in two variations. First is the additive variant; it is suitable when seasonal changes are stable over time. Second is the multiplicative variant, when seasonal fluctuations are not stable. Holt-Winter performs better for short-term forecasting. The Holt-Winters model incorporates a smoothing equation for level, trend, and seasonality. By incorporating these elements, Holt-Winters can perform better short-term forecasts. Level is the current baseline value, trend represents the slope, the rate at which data is increasing or decreasing over time, and trend shows the repeating patterns. This model is applicable in various domains to forecast demand and seasonal variations.

4 Results and Discussion

This study is focused on building and identifying machine learning models suitable for the agriculture dataset, which is downloaded from the Department of Statistics in Jordan. This machine learning models' goal is to forecast the crop yield so that all the stakeholders of the agriculture cycle can take informed decisions. The total planted area for different crops is compared with the harvested area to understand the productivity of the different crops. Figure 3 shows the difference in planted and harvested areas. Trends in plantation and harvesting from 1999 to 2022 show a significant decline. Annual crop yield is measured in metric tons; among all the crops, barley and wheat show the highest production in Jordan. Figure 2 illustrates the annual production of all crops, and it indicates a decline in the annual production of crops. Descriptive analytics in this study highlights the difference in planted area and harvested area, declining trends, and decrease in annual production rate. An exploratory data analysis is conducted on the given dataset to visualize patterns and correlations among multiple agricultural variables. Table 2 shows the correlation between production and independent variables; based on the correlations, it became clear that the planted and harvested areas impact production. Multiple algorithms are used to build models, and the performance of each model is measured on a scale of R-squared value and error rate. The error metrics used here are mean squared error (MSE), root mean squared error (RMSE), R-squared errors, and mean absolute percent error (MAPE). Error metrics are shown in Table 3; the multiple linear regression model attained an R-squared value of 0.994 and an MSE of 0.024 square metric tons (t^2). Lasso regression also outperforms other models, showing the high R-squared value and mean squared error of 0.023. Lasso regression even outperforms multiple linear regression on error matrices. Polynomial regression shows the mean squared error of 0.238, RMSE of 0.296, and R-squared value of 0.981.

Ridge Regression shows the MSE of 0.035, which is slightly higher than linear and Lasso, and the R-squared value of 0.992, which is very high, but slightly lower than linear and Lasso. Ridge regression achieves strong performance, comparable to linear and Lasso regression. Its L2 regularization effectively reduces multicollinearity without being as aggressive in feature selection as Lasso. ElasticNet Regression performs the worst among all the models, with a high error rate and low R-squared values. MSE of 1.308 and R-square of 0.725. The gap between the lower adjusted R-squared and the high error metrics could indicate problems with the regularization parameters or that the model might not be well-suited for this dataset. SVM Regression shows moderate performance with an MSE of 0.271, considerably higher than linear and tree-based models, and an adjusted R-squared of 0.996, which is very high. Decision Tree Regression model performance is not significant among the ensemble tree methods, with an MSE of 0.243, which is considerably higher than linear and ensemble tree methods. Although the adjusted R-squared is 0.996, which is very high.

Random Forest results show an MSE of 0.023, the lowest among all models, and an R-squared value of 0.993 and an adjusted R-squared value of 0.999, the highest among all models. Random Forest shows outstanding performance, rivaling or even surpassing Lasso and Multiple Linear Regression. As an ensemble method, it likely excels by minimizing overfitting and effectively capturing complex relationships within the data, as indicated by the Out-of-Bag (OOB) score of 0.733.

The XGBoost model prediction is 98% with an R-squared error of 0.98, which explains the variance in the model, and it captures the factors influencing the production. The average squared error rate is 0.092, which is also significantly low. Overall, this model performance is lower than multilinear models and random forest and lasso regression, which show excellent performance. The mean absolute percentage error of multilinear regression is 1.27%, random forest is 1.57%, and XGBoost is 3.25%.

Findings in this study show the importance of choosing an appropriate model for the data. Figure 4 presents a comparison of the MSE across all models; it describes low mean squared errors of 0.024 for multilinear regression and 0.023 for lasso regression and random forest. While Figure 5 illustrates the R-squared scores, which explain variance in the dependent variables based on independent variables. The R-squared score of multiple linear regression is 0.994, ridge regression 0.992, lasso regression 0.995, and random forest regression 0.993. These results suggest that multiple machine learning models can effectively perform on the given agriculture dataset.

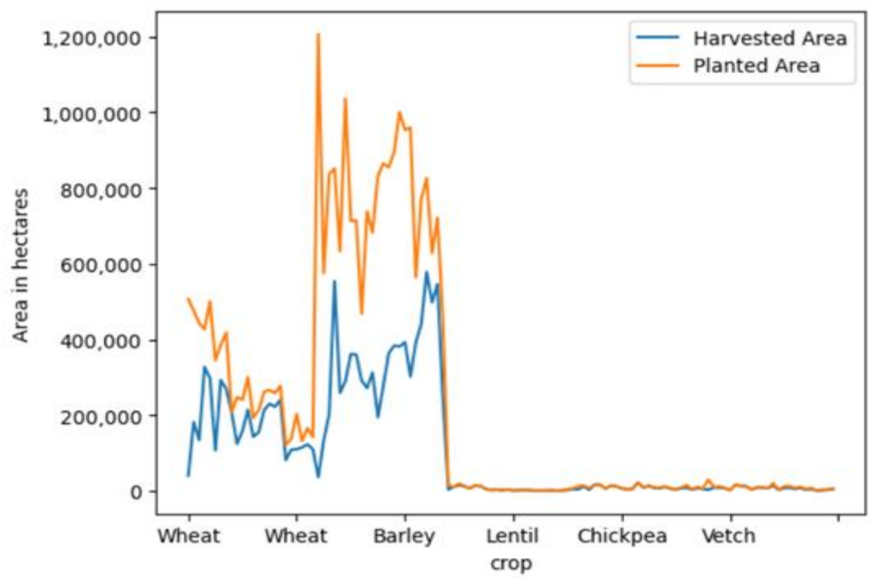

Figure 3: Comparison of planted area and harvested area

Table 3: Result of multiple models

Method	MSE	MAE	MAPE	R-squared	RMSE	Adjusted R-squared Value
Multiple linear regression	0.024	0.106	1.27%	0.994	0.156	0.998
Lasso regression	0.023	0.104	1.23%	0.995	0.153	0.999
Polynomial Regression	0.238	0.087	3.34%	0.981	0.296	0.998
Ridge Regression	0.035	0.126	1.45%	0.992	0.187	0.999
ElasticNet Regression	1.308	0.995	13.73%	0.725	1.143	0.982
SVM Regression	0.271	0.368	4.66%	0.943	0.521	0.996
Decision Tree Regression	0.243	0.368	5.19%	0.948	0.493	0.996
Random forest	0.023	0.108	1.57%	0.993	0.153	0.999
XGBoost	0.092	0.236	3.25%	0.980	0.303	0.998

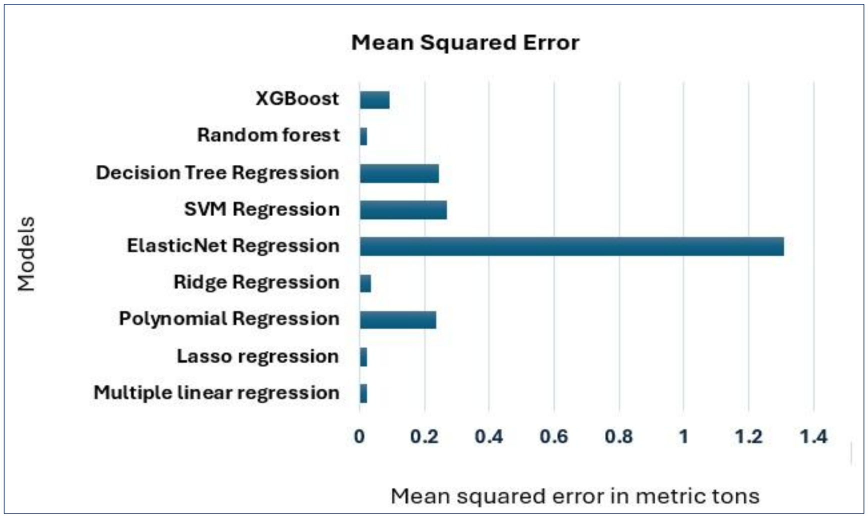

Figure 4: Mean squared error for the applied algorithms

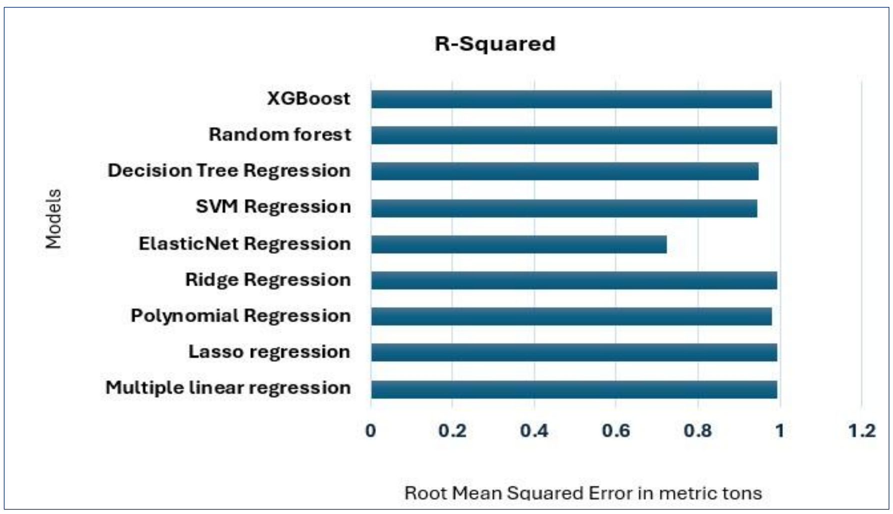

Figure 5: R-squared error for the applied algorithms

Future scenario predictions are tested using dependent variable data generated through the moving averages method, utilizing the average of the past five years. The dataset includes data up to 2022. The columns for planted area, harvested area, and average yield data from 2023 to 2032 are generated using moving averages, with temperature data also produced by the same method. The time series methods predict future production using ARIMA and Holt-Winters models (Agarwal & Tarar, 2021). Crop production from 2023 to 2032 has been predicted, with the results presented in Table 4. Crop yield prediction using ARIMA and Holt-Winters models is shown in Figures 6 and 7. Crop production forecasts are primarily based on in-season variables, statistical regressions between historical yields, and field surveys. While field surveys remain the predominant method for forecasting crop yields in most countries, statistical regression using historical data is also significant in predicting crop yields, and machine learning allows computer systems to learn patterns from data using statistical methods, without writing specific instructions for every task. Table 5 presents the security evaluation metrics based on the regression results for different models (see Table 3); these metrics enable us to determine the effectiveness of the smart agriculture data/IoT-enabled system for threat detection and event management of security situations.

Table 4: Prediction of crop production with adaptive forecasting models

Year	Wheat		Barley		Lentil		Chickpea		Vetch	
	ARIMA	Holt	ARIMA	Holt	ARIMA	Holt	ARIMA	Holt	ARIMA	Holt
2023	19898.9	17284.3	32339.0	48638.7	140.0	539.9	3960.7	3432.4	568.9	673.3
2024	20857.9	16911.1	30352.3	48547.5	358.5	519.7	3402.6	3530.5	186.9	684.8
2025	20749.9	16537.8	30176.3	48456.2	642.3	499.4	3318.7	3628.7	447.9	696.2
2026	21849.9	16164.5	44923.2	48364.9	629.2	479.2	4044.9	3726.8	731.4	707.7
2027	22181.4	15791.2	65106.2	48273.7	426.7	459.0	3769.4	3824.9	689.2	719.1
2028	21557.7	15418.0	57286.8	48182.4	314.4	438.8	4018.7	3923.0	692.9	730.6
2029	21668.4	15044.7	52714.4	48091.1	446.9	418.6	3699.7	4021.2	434.5	742.0
2030	21514.6	14671.4	48884.0	47999.9	597.4	398.3	3756.8	4119.3	398.0	753.5
2031	21342.6	14298.1	40473.2	47908.6	535.5	378.1	3751.0	4217.4	535.6	765.0
2032	21597.3	13924.9	35308.1	47817.3	397.2	357.9	3888.3	4315.6	606.9	776.4

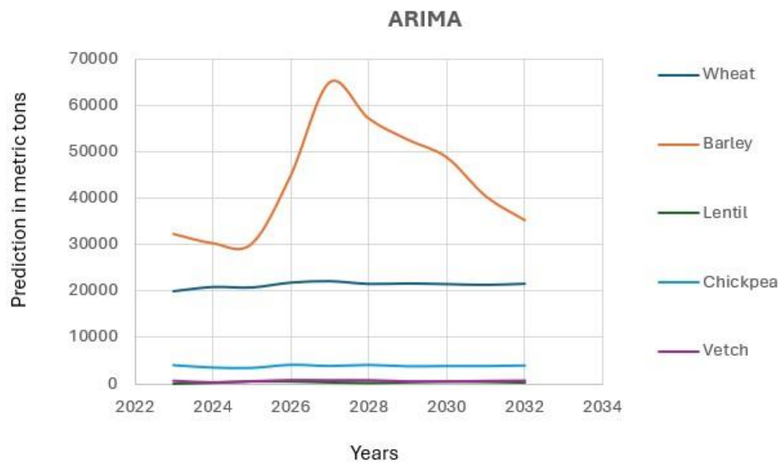

Figure 6: Prediction of crop yield with ARIMA (2023-2032)

The deseasonalized values indicate a downward tendency over time. Commencing at a peak of 26,225.97 in 1999, the deseasonalized numbers predominantly decline, signifying that production is heading downward even after accounting for seasonal variations. The declining trend line from approximately 26,225.97 in 1999 to a much lower production rate till 2022 indicates a substantial decline. Forecasts for the period from 2023 to 2032 are presented in Figure 8. Forecasts exhibit a steady rise, starting at an expected 19,898.88 metric tons in 2023 and ending at 21,597.28 by 2032. Forecasting shows an anticipated recovery in coming years; ARIMA and Holt-Winter model predictions show growth in crop yield. This recovery in production rate shown by forecasting models is possible by a reversal or upturn not recorded in historical data.

Table 5: Security evaluation metrics based on the regression performance models

Model	TPR (Recall)	Precision	FPR	F1-Score	AUC-ROC	MTTD (s)	MTTR (s)
Multiple Linear Regression	91.5%	89.2%	4.8%	90.3%	0.92	18	42
Lasso Regression	92.1%	90.0%	4.5%	91.0%	0.94	17	40
Polynomial Regression	88.3%	85.7%	6.9%	87.0%	0.89	22	50
Ridge Regression	90.7%	88.5%	5.2%	89.6%	0.91	19	45
ElasticNet Regression	79.2%	76.3%	11.2%	77.7%	0.81	28	63
SVM Regression	86.8%	84.0%	7.5%	85.4%	0.88	23	53
Decision Tree Regression	87.6%	85.1%	7.1%	86.3%	0.90	21	49
Random Forest	92.0%	90.5%	4.3%	91.2%	0.95	16	38
XGBoost	89.4%	87.9%	6.1%	88.6%	0.92	20	47

TPR: True Positive Rate, FPR: False Positive Rate, F1-Score: balance of precision, AUC-ROC: ability to discriminate attack vs normal across thresholds, MTTD: Mean Time to Detect, MTTR: Mean Time to Respond.

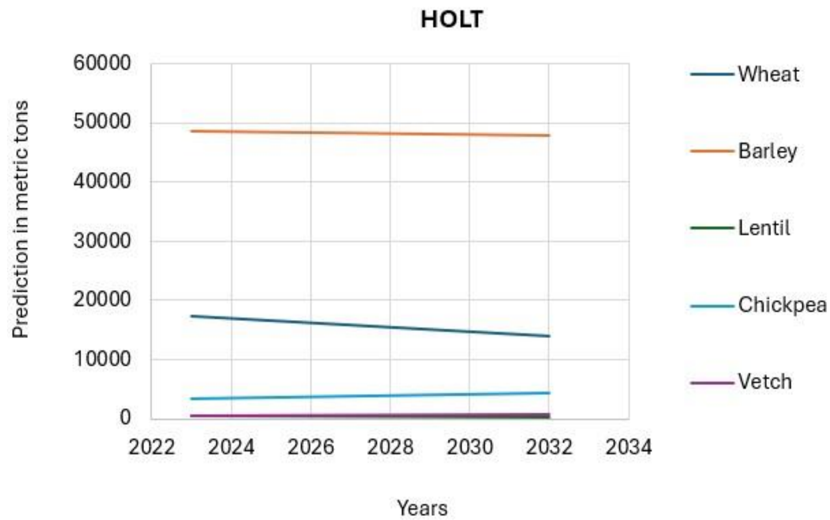

Figure 7: Holt-winters prediction of crop yield

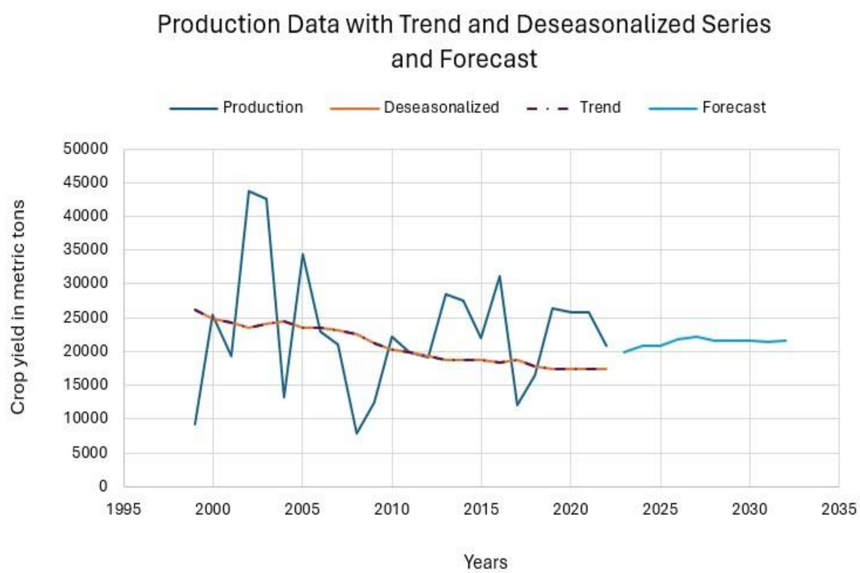

Figure 8: Projected values for the prediction of crop yield from 2023 to 2032

5 Conclusion and Future Work

This study builds multiple machine learning models to predict the crop yield and descriptive analytics to show the underlying factors and trends in historical data. Descriptive analytics illustrates the difference between planted area and harvested area. Among the nine machine learning models, random forest regression, multilinear regression, and lasso regression predict with an R-squared value of 0.99 with mean absolute percent errors of 1.57%, 1.27%, and 1.23%, respectively. This study finds the potential of machine learning models in predicting the crop yield with high R-squared value; some of the models best fit the given agricultural data. For a dataset with a non-linear relationship between dependent and independent variables, some of the models as discussed in the results section are suitable for this dataset, which encapsulates the variance to forecast with a high R-squared value and low error rates. Findings also highlight the fact that all of the models do not perform well. Overfitting and underfitting challenges are also considered in performance analysis. A limitation of this study is that having more data variables to represent other external factors could result in more reliable model building. With existing historical data and by an ensemble of machine learning methods, this research contributes to building a more resilient and accurate prediction framework. This type of study presents the advantages of incorporating machine learning and time series forecasting models in the efficient prediction of crop yield. Adoption of these techniques supports food security by providing farmers with advanced tools to make informed decisions and mitigate the threats of climate change and pest control and adopt robust and resilient logistical and crop management practices. Future work can use IoT and remote sensing technologies in capturing the real-time soil and crop data to increase the reliability and short-term decision-making for farmers. More advanced machine learning models, deep learning models like recurrent neural networks and convolutional neural networks, will open new frontiers in agricultural forecasting.

In conclusion, the results from this study can support farmers, researchers and policy makers in understanding the opportunities and challenges in agriculture. Encourages all parties to adopt smart agricultural technologies in crop yield forecasting.

Informed Consent Statement: “Not applicable”.

Acknowledgment: “The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/26504)”.

Data Availability Statement: “The Dataset was uploaded within supplementary files via the submission system”.

Conflicts of Interest: “The authors declare no conflicts of interest.”

References

[1] Agarwal, S., & Tarar, S. (2021). A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1714, No. 1, p. 012012). IOP Publishing. https://doi.org/10.1088/1742-6596/1714/1/012012

[2] Albishi, S., Soh, B., Ullah, A., & Algarni, F. (2017). Challenges and Solutions for Applications and Technologies in the Internet of Things. *Procedia Computer Science*, *124*, 608-614. https://doi.org/10.1016/j.procs.2017.12.196

[3] Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE access*, *7*, 129551-129583. https://doi.org/10.1109/ACCESS.2019.2932609

[4] Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *Advances in agronomy*, *154*, 201-255. https://doi.org/10.1016/bs.agron.2018.11.002

[5] Behera, S. K., Sethy, P. K., Sahoo, S. K., Panigrahi, S., & Rajpoot, S. C. (2021). On-tree fruit monitoring system using IoT and image analysis. *Concurrent Engineering*, *29*(1), 6-15. https://doi.org/10.1177/1063293X20988395

[6] Cha, S. C., Hsu, T. Y., Xiang, Y., & Yeh, K. H. (2018). Privacy enhancing technologies in the Internet of Things: Perspectives and challenges. *IEEE Internet of Things Journal*, *6*(2), 2159-2187. https://doi.org/10.1109/JIOT.2018.2878658

[7] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

[8] Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE access*, *8*, 86886-86901. https://doi.org/10.1109/ACCESS.2020.2992480

[9] Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, *13*(16), 9288. https://doi.org/10.3390/app13169288

[10] Ferrag, M. A., Maglaras, L. A., Janicke, H., Jiang, J., & Shu, L. (2017). Authentication protocols for internet of things: a comprehensive survey. *Security and Communication Networks*, *2017*(1), 6562953. https://doi.org/10.1155/2017/6562953

[11] Gladence, L. M., Reddy, K. R., Reddy, M. P., & Selvan, M. P. (2021, June). A prediction of crop yield using machine learning algorithm. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1072-1077). IEEE.

[12] Gupta, S., Geetha, A., Sankaran, K. S., Zamani, A. S., Ritonga, M., Raj, R., ... & Mohammed, H. S. (2022). Machine Learning-and Feature Selection-Enabled Framework for Accurate Crop Yield Prediction. *Journal of Food Quality*, *2022*(1), 6293985. https://doi.org/10.1155/2022/6293985

[13] Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction. https://doi.org/10.1007/978-0-387-84858-7_3

[14] Hukare, V., & Kumbhar, V. (2025). Optimization of feature selection methods to improve the performance of machine learning models for crop yield prediction. *ES Food and Agroforestry*, *20*, 1474. http://dx.doi.org/10.30919/faf1474

[15] Jorvekar, P. P., Wagh, S. K., & Prasad, J. R. (2024). Predictive modeling of crop yields: A comparative analysis of regression techniques for agricultural yield prediction. *Agricultural Engineering International: CIGR Journal*, *26*(2).

[16] Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 124-135. https://doi.org/10.1016/j.isprsjprs.2019.11.008

[17] Kaya, Y., & Polat, N. (2023). A linear approach for wheat yield prediction by using different spectral vegetation indices. *International Journal of Engineering and Geosciences*, *8*(1), 52-62. https://doi.org/10.26833/ijeg.1035037

[18] Lad, A. M., Bharathi, K. M., Saravanan, B. A., & Karthik, R. (2022). Factors affecting agriculture and estimation of crop yield using supervised learning algorithms. *Materials Today: Proceedings*, *62*, 4629-4634. https://doi.org/10.1016/j.matpr.2022.03.080

[19] Liliane, T. N., & Charles, M. S. (2020). Factors affecting yield of crops. *Agronomy-climate change & food security*, *9*, 9-24. https://doi.org/10.5772/intechopen.90672

[20] Mark, R. (2019). Ethics of using AI and big data in agriculture: The case of a large agriculture multinational. *The ORBIT Journal*, *2*(2), 1-27. https://doi.org/10.29297/orbit.v2i2.109

[21] Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010. https://doi.org/10.1016/j.ailsci.2021.100010

[22] Mukherjee, S., Mishra, A., & Trenberth, K. E. (2018). Climate change and drought: a perspective on drought indices. *Current climate change reports*, *4*(2), 145-163. https://doi.org/10.1007/s40641-018-0098-x

[23] Nusir, M. (2024). Research Data. Agriculture Jordan. https://drive.google.com/drive/folders/1t_FvKyPJnpjk0PUVWrmjI2mpUTf1Rh0t?dmr=1&ec=wgc-drive-globalnav-goto.

[24] Páez, A., & Boisjoly, G. (2023). Exploratory data analysis. In *Discrete choice analysis with R* (pp. 25-64). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-20719-8_2
[25] Panigrahi, B., Kathala, K. C. R., & Sujatha, M. (2023). A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. *Procedia Computer Science*, *218*, 2684-2693. https://doi.org/10.1016/j.procs.2023.01.241
[26] Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, *37*(5), 2671-2692. https://doi.org/10.1007/s00180-022-01207-6
[27] Patil, Y., Ramachandran, H., Sundararajan, S., & Srideviponmalar, P. (2025). Comparative analysis of machine learning models for crop yield prediction across multiple crop types. *SN Computer Science*, *6*(1), 64. https://doi.org/10.1007/s42979-024-03602-w
[28] Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*, 103016. https://doi.org/10.1016/j.agsy.2020.103016
[29] Rahman, M., Islam, M. M., Kim, H. J., Alam, M., Sadiq, S., Rahman, M. K., ... & Dewan, A. (2023). Integrated Deep Learning and Genetic Algorithm Approach for Groundwater Potential Zone Prediction Incorporating Cmip6 Gcm: Unveiling Synergies for Enhanced Water Resource Management. https://dx.doi.org/10.2139/ssrn.4581333
[30] Rifna, E. J., Dwivedi, M., Seth, D., Pradhan, R. C., Sarangi, P. K., & Tiwari, B. K. (2024). Transforming the potential of renewable food waste biomass towards food security and supply sustainability. *Sustainable Chemistry and Pharmacy*, *38*, 101515. https://doi.org/10.1016/j.scp.2024.101515
[31] Samal, S., Acharya, B., & Barik, P. K. (2022). Internet of Things (IoT) in agriculture toward urban greening. In *AI, Edge and IoT-based Smart Agriculture* (pp. 171-182). Academic Press. https://doi.org/10.1016/B978-0-12-823694-9.00015-3
[32] Sidhu, B. S., Mehrabi, Z., Ramankutty, N., & Kandlikar, M. (2023). How can machine learning help in understanding the impact of climate change on crop yields?. *Environmental Research Letters*, *18*(2), 024008. https://doi.org/10.1088/1748-9326/acb164
[33] Singh, R., Das, R., Sangwan, S., Rohatgi, B., Khanam, R., Peera, S. P. G., ... & Misra, S. (2021). Utilisation of agro-industrial waste for sustainable green production: a review. *Environmental Sustainability*, *4*(4), 619-636. https://doi.org/10.1007/s42398-021-00200-x
[34] Xu, J., Gu, B., & Tian, G. (2022). Review of agricultural IoT technology. *Artificial Intelligence in Agriculture*, *6*, 10-22. https://doi.org/10.1016/j.aiia.2022.01.001
[35] Zhang, H., Wei, X., Zou, T., Li, Z., & Yang, G. (2014, September). Agriculture Big Data: Research status, challenges and countermeasures. In *International conference on computer and computing technologies in agriculture* (pp. 137-143). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19620-6_17

Authors Biography

Muneer Nusir I was awarded my PhD degree from Brunel University-London (England) in July 2015, I have been Assistant professor (Feb 2016-present) at the Information Systems Department of Prince Sattam bin Abdulaziz University, Kingdom of Saudi Arabia. I attended a number of training courses and member of the IEEE-branch of Alkharj. I have been publishing several research Papers (International Conferences & refereed Journals). My research interest falls within Digital Service Co-design, Human centered Design, Business Intelligence, Health Informatics, Smart city and business process variability management. I have teaching, practical and research experience. My experience in teaching incorporates with different universities in Jordan, UK and Saudi Arabia.

Mohammad Alshirah Ph.D., is an Associate Professor at the Information Systems Department at Al al-Bayt University. He received his BSc. in Software Engineering from the Hashemite University, Jordan in 2007 and received his MSc in Computer Science from Al al-Bayt University, Jordan in 2010. He obtained his PhD in Software Engineering at the Department of Computer Science of the University of Leicester in the United Kingdom in 2016. He has a variety of academic and professional qualifications and experience. His research interests include, but are not limited to, Machine Learning, Software Usability and User Experience.

Sahar Al Mashaqbeh received a B.S.c degree in Mechatronics engineering and an M.S.c degree in Industrial Engineering from the Hashemite University in 2006 and 2011, respectively, and another M.S.c in quality improvement engineering and Ph.D. degrees in industrial engineering from the University of Bradford, U.K., in 2019 and 2020, respectively. She currently serves as an assistant professor at the Hashemite University, Jordan. Furthermore, since November 2021, she has been working as an Honorary Visiting Research Fellow with the University of Bradford.

Rayeh Alghsoon is a dedicated computer engineer and educator with experience in teaching, research, and project coordination. She currently assists AARINENA's Executive Secretary in advancing agricultural research in the Near East and North Africa, leveraging her organizational skills and technical expertise. Raiah has taught various engineering courses at Al-Ahliyya Amman University and has contributed to research initiatives at the Jordan University of Science and Technology. She holds an MSc in Computer Engineering and Networks, with published research in smart cities, AMOLED displays, and nanotechnology.)